DOCUMENT RESUME

ED 381 567 TM 022 927

TITLE	College-Level Academic Skills Test, Technical Report, 1989-90.
INSTITUTION	Florida State Dept. of Education, Tallahassee.
PUB DATE	90
NOTE	53p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Academic Achievement; Accountability; Achievement Tests; Associate Degrees; *College Students; *Communication Skills; *Degree Requirements; Higher Education; Language Skills; *Mathematics Achievement; Public Colleges; Racial Differences; Reading Achievement; Scores; Scoring; Sex Differences; State Legislation; Test Construction; *Test Results
IDENTIFIERS	*College Level Academic Skills Test; *Florida

ABSTRACT

The College-Level Academic Skills Test (CLAST) is part of Florida's system of educational accountability that is mandated by state law. The CLAST is an achievement test measuring students' attainment of college-level communication and mathematics skills identified by faculties of community colleges and state universities. Since August 1, 1984 students in public institutions in Florida have been required to have CLAST scores that satisfy state standards for the award of an associate in arts degree and for admission to upper division status in a state university in Florida. In addition, students in private schools may need CLAST scores to receive state financial aid. The CLAST consists of essay, English language skills, reading, and mathematics tests. Test development is traced, and the test itself is described, along with scoring and development information. Summary data are presented for first-time takers in 1989-90. Passing rates are presented for groups of students classified by race/ethnicity and gender, as well as college status. Fourteen tables present test data for the 1989-90 school year. Six appendixes describe the test in greater detail and list College-Level Academic Skills Project (CLASP) and state-level task force members, 1989-90. (Contains 15 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

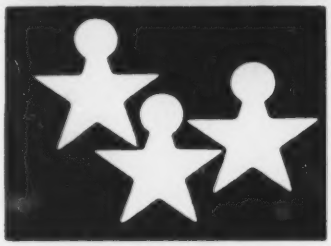
ROBERT BEDFORD

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED 381 567

CLAST

COLLEGE-LEVEL ACADEMIC SKILLS TEST



TECHNICAL REPORT

1989-90

State of Florida
Department of Education
College-Level Academic Skills Project
Tallahassee, Florida
Betty Castor, Commissioner
Affirmative action/equal opportunity employer

BEST COPY AVAILABLE

TM022927

CLAST TECHNICAL REPORT, 1989-90

CONTENTS

I. OVERVIEW (1)

Eligibility to Take the CLAST (1)
Test Administration Plan (2)

II. DEVELOPMENT OF THE CLAST (3)

Background (3)
Identification of Skills (3)
Review of Available Tests (4)
Development of Test Specifications (5)
Development of Item Specifications (5)
Development of Items (6)
Development of CLAST Standards (7)

III. DESCRIPTION OF THE CLAST (9)

Test Specifications (9)
Item Bank (9)
Test Assembly (9)
Test Instructions (10)
Quality Control (11)

IV. TECHNICAL CHARACTERISTICS OF THE CLAST (13)

Test Score Equating (13)
Reliability of Scores (15)
Item Analysis (18)
Preventing Item Bias (20)
Validity of Scores (20)

V. SCORING AND REPORTING PROCEDURES (23)

Scoring Activities (23)
Score Scales (23)
Essay Scoring (27)
Reporting Test Results (29)
Interpreting and Using Scores (32)

VI. SUMMARY OF 1989-90 RESULTS (33)

BIBLIOGRAPHY (41)

APPENDICES (43)

TABLES

Table

1. Communication and Computation Competencies and Broad Skills (4)
2. Standards (Passing Scores) for CLAST Subtests (7)
3. CLAST Specifications by Subtest, 1989-90 (10)
4. Multiple-Choice Raw Score Reliability Statistics, 1989-90 (16)
5. Summary Data for All Essay Readers, 1989-90 (17)
6. Essay Reader Agreement after Referee, 1989-90 (17)
7. Alpha Coefficients, 1989-90 (18)
8. English Language Skills Score Conversions, 1989-90 (24)
9. Reading Score Conversions, 1989-90 (25)
10. Mathematics Score Conversions, 1989-90 (26)
11. Raw and Scaled Scores, 1989-90 (34)
12. Percentage of Examinees Passing All Four Subtests, 1989-90 (35)
13. Essay Mean Scaled Scores, 1989-90 (36)
14. English Language Skills Mean Scaled Scores, 1989-90 (37)
15. Reading Mean Scaled Scores, 1989-90 (38)
16. Mathematics Mean Scaled Scores, 1989-90 (39)

I. OVERVIEW

The College-Level Academic Skills Test (CLAST) is part of Florida's system of educational accountability and is mandated by Section 229.551(3)(k), FS, 1986. The CLAST is an achievement test measuring students' attainment of college-level communication and mathematics skills identified by faculties of community colleges and state universities through the College-Level Academic Skills Project (CLASP). The skills (Appendix A) have been adopted by the State Board of Education (SBE) through Rule 6A-10.0310, FAC. Provisions for keeping the skills list current, maintaining active participation of faculty members in the implementation of the testing program, and administering the test are provided in the CLAST Test Administration Plan.

The CLAST consists of four subtests: essay, English language skills, reading, and mathematics. Each subtest yields a single score reported to the student and to the institution needing the scores. Students also receive broad skill information useful in identifying areas of possible strength or weakness. While the CLAST does not yield the skill-by-skill information necessary for full diagnosis of individual student needs, institutions can identify areas of need for groups of students by aggregating scores into broad skills over several administrations. Although CLAST scores relate positively to other measures of academic performance, they do not predict examinees' future performance in upper division programs.

Since August 1, 1984, students in public institutions in Florida have been required to have CLAST scores which satisfy the standards set forth in SBE Rule 6A-10.0312, FAC, for the award of an associate in arts degree and for admission to upper division status in a state university in Florida. In addition, students in private institutions may need CLAST scores to receive state financial aid.

Statutes and rules pertaining to the CLAST requirement are contained in the CLAST Test Administration Plan.

Eligibility to Take the CLAST

The CLAST may be taken by any student who is seeking an associate in arts or baccalaureate degree and who applies to take the test by the deadline established for registration. Students who have previously taken the CLAST and have not passed all subtests may apply at any regular administration to retake the subtest(s) not passed.

In addition, participating colleges and universities are to register other students who meet both of the following criteria:

1. The students are eligible to participate in a State of Florida financial aid program governed by SBE Rule 6A-20.005, FAC.

2. The students are required under provisions of SBE Rule 6A-20.005, FAC, to have CLAST scores to continue their eligibility beyond the academic term in which they register for the CLAST.

BEST COPY AVAILABLE

Although CLAST scores are not needed to receive an associate in science degree, students who are in that program may be registered for the CLAST if they satisfy the requirements for (1) the associate in arts degree or (2) admission to upper division status.

In all cases, registration of students for the CLAST must be made in an institution which can determine the eligibility of applicants to take the test. Thus, registration normally will be done by the institution in which students are enrolled during the term in which they will take the test. However, an applicant for upper division status in a state university who needs CLAST scores and meets eligibility requirements, but is not enrolled in an institution which administers the CLAST, may be registered for the test in the institution that needs the scores.

Students must apply to take the test on or before the registration deadline established for that administration.

Test Administration Plan

Under provisions of Section 229.551(3)(k), Florida Statutes, the Commissioner of Education maintains statewide responsibility for the administration of the CLAST.

A plan for the administration of the CLAST for the 1989-90 academic year was issued by the Commissioner in September 1989. The plan, developed by the Department of Education, assigns administrative responsibility for the CLAST at three levels: the Department of Education; the Technical Support Contractor; and the community colleges and state universities which administer the test to eligible students. The Office of Instructional Resources of the University of Florida is the Technical Support Contractor.

The plan also describes the policies and procedures under which the testing program operates. The CLAST Test Administration Manual and the CLAST Institutional Test Administrator's Manual, which are made a part of the plan, give additional specific information to assist institutional personnel in carrying out their responsibilities.

II. DEVELOPMENT OF THE CLAST

The test development process for the CLAST began with identifying skills to be assessed and continues with developing items for inclusion in the test. This chapter describes the major developmental efforts culminating in the first test administration, the item development procedures, and the development of standards (passing scores).

Background

In 1979 the Florida Legislature, through Florida Statute 79-222 (now Section 229.551), enacted legislation requiring the identification of skills to measure the achievement of essential academic skills of college students. The Department of Education then charged the Articulation Coordinating Committee with the task of implementing that part of the legislation dealing with the identification of skills and tests to measure achievement of those skills. The result was the Essential Academic Skills Project (EASP, now CLASP). The EASP included an executive committee, a project director, a state-level task force on communication, a state-level task force on computation, and a state-level standing committee on student achievement. Current members are identified in Appendix B.

Identification of Skills

The state-level task forces, together with the project director and other project personnel acting in an advisory capacity, worked to identify essential academic skills which every student, regardless of major, should have acquired by the end of the sophomore year. The task forces worked through a series of meetings from January to November of 1980 with input from institutional-level task forces which had been established to involve faculty members in Florida's public universities and community colleges in the identification of the skills.

The task forces identified four generic competencies (reading, listening, writing, and speaking) in communication and four generic competencies (algorithms, concepts, generalizations, and problem solving) in computation. Each generic competency was subsequently reviewed and broad skill categories developed for each competency.

Skills were then developed for each broad skill category. These skills were presented to a random sample of faculty members from broad discipline areas in Florida's public community colleges and universities. Based on the results of the survey, the task forces made recommendations to the SBE. In September 1981 the SBE adopted all of the skills recommended by the task forces. During 1985 and 1989, an extensive review of the CLASP skills resulted in the addition, deletion, and/or modification of some of the original skills. As a result of the 1985 review, revised skills were adopted by the SBE and have been measured by the CLAST since the fall 1987 administration (table 1); the revised skills resulting from the 1989 review will be incorporated into the CLAST with the fall 1992 administration.

TABLE 1

Communication and Computation Competencies and Broad Skills

Communication	Computation
Reading	Algorithms
Literal Comprehension	Arithmetic
Critical Comprehension	Geometry and Measurement
	Algebra
Listening	Statistics, including Probability
Literal Comprehension	Logical Reasoning
Critical Comprehension	
	Concepts
Writing	Arithmetic
	Geometry and Measurement
Multiple-Choice	Algebra
Word Choice	Statistics, including Probability
Sentence Structure	Logical Reasoning
Grammar, Spelling, Capitalization, and Punctuation	
	Generalization
Essay	Arithmetic
Suitability to Purpose and Audience	Geometry and Measurement
Effectiveness and Conformity to Standard English	Algebra
	Statistics, including Probability
Speaking	Logical Reasoning
Composition of Message	
Transmission of Message	Problem Solving
	Arithmetic
	Geometry and Measurement
	Algebra
	Statistics, including Probability
	Logical Reasoning

Review of Available Tests

Once the skills had been identified, the Standing Committee on Student Achievement, with the assistance of project staff, began its task of identifying tests and other assessment procedures which could be used to measure achievement of the skills. To accomplish the task, an extensive search was conducted to review commercially available tests and tests developed by community colleges and state universities which might be appropriate for measuring achievement of communication and computation skills. Sixty-six communication tests and fifty-four computation tests were reviewed in depth. Though all of the tests addressed some of the skills, none was judged adequate for measuring all of the skills identified in SBE Rule 6A-10.0310, FAC.

It was recommended that three multiple-choice subtests be developed in the areas of writing, reading, and computation. Since all of the writing

skills could not be tested using a multiple-choice format, it was further recommended that an essay test be developed to measure the entire set of writing skills. Although students should have obtained the listening and speaking skills by the time they complete their sophomore year in college, no statewide tests had been developed to measure student achievement of those skills.

A more detailed report on the test search may be found in Test Search and Screen for College-Level Communication and Computation Skills (Department of Education, May 1981).

Development of Test Specifications

Specifications for a test which could be used to measure the achievement of the skills listed in SBE Rule 6A-10.0310, FAC, were developed between April and August of 1981 by the project director and staff, with assistance from the Standing Committee on Student Achievement, the communication and computation task forces, and measurement consultants. Recommendations of state-level task force members about the assessment of the skills, as well as practical and measurement issues, were considered in determining the nature of the subtests and the number of items to be included in each subtest. These same procedures were followed for revising the test specifications necessitated by the 1985 and 1989 skill revisions. Specifications for the 1989-90 forms are described in Chapter III.

Development of Item Specifications

After test specifications were developed, formulation of item specifications began. During the fall of 1981, item specifications were written for the reading and writing skills, as well as the computation skills dealing with algorithms and concepts. In 1983 item specifications for computation skills dealing with generalizations and problem solving were written and reviewed. Concurrently, the original specifications for the essay, writing, and reading items were reviewed again and revised as necessary. This process was repeated following the 1985 and 1989 revisions.

All specifications were written by the chairpersons of the state-level task forces with assistance from task force members, standing committee members, content and measurement consultants, and Department of Education staff. Reviews of the specifications were done by faculty members from community colleges and state universities. Appendix D lists the 1989 review team.

Item writers used the item specifications as guides for item content and format. Copies of item specifications were distributed for use in all thirty-seven community colleges and state universities to aid faculties in planning for instruction and assessment of the skills. Copies of item specifications are available in the institutions as well as from the Department of Education.

Development of Items

Items are developed for the CLAST through contracts with post-secondary faculty who write, review, pilot test, and revise items based on item specifications and recommendations of state-level item review committees. Items developed under these contracts are submitted to the Department of Education for field-testing and analysis. The following procedures are used to develop and approve test items for the CLAST.

1. A contractor is selected based on its qualifications, including its past performance as an item developer and the qualifications of its item writers and reviewers.

2. The contractor holds a training session for item writers and reviewers to discuss test security issues, purpose of the CLAST, use of item specifications, characteristics of good test items, item bias issues, and specific assignments to the contractor.

3. Initial drafts of items are written and reviewed by members of the contractor's item writing team.

4. Items are pilot tested with college students, and the results of the pilot test and suggestions from other item writers are used in revising the items. The pilot test involves administering each item to about thirty students and interviewing at least five of them to obtain specific information about the items.

5. Based on pilot test data, items are reviewed and revised by members of the contractor's review team who have not been involved in the item writing. Attention is given to content, measurement, and bias issues (Appendix C).

6. Revised items are submitted to the Department of Education, and a state-level committee is convened to review the items and recommend revisions and/or deletions in the contractor's set.

7. Based on state-level review, items are revised by the contractor's team and submitted to the Department of Education in final form.

8. Items are then included in the CLAST as developmental items and are not counted as scored items for students. This produces classical and Rasch item statistics for evaluating item quality. Items are screened based on the following criteria: p-value greater than or equal to .40, point-biserial greater than or equal to .30, Rasch fit between less than or equal to 3.0, and Rasch total fit less than or equal to 1.0 + 3 standard errors. These criteria represent an ideal level of functioning for an item. If the item point-biserial statistic is less than 0.30, the item may still be considered for use on a future examination if it measures an important dimension of a required objective. Items are not used if the point biserial correlation coefficients are close to or less than zero.

9. Essay topics are field-tested by a qualified contractor. Data generated for topic evaluations include distribution of scores, number of essays written, number written off topic, mean score, median score,

percentage of complete agreement between raters, percentage of agreement within one score point, alpha coefficients with and without referee, and reader comments. Topics are evaluated in terms of clarity, relevance and appeal to the target population, suitability for development of an essay, and potential biasing elements. The contractor recommends the topics suitable for inclusion in the CLAST and identifies any potential problems.

In 1989-90 the Department of Education awarded a grant to Miami-Dade Community College to fieldtest previously developed essay topics for future forms of the CLAST and to the Office of Instructional Resources at the University of Florida to score the resultant essays.

Development of CLAST Standards

CLAST standards (passing scores) were set by the SBE in March 1984. The passing scores reflected the judgment of a state-level panel of interested persons concerning the minimum level of performance acceptable for the successful completion of the sophomore year in community colleges and state universities in Florida. SBE Rule 6A-10.0312(1), FAC, establishes minimum standards, in terms of scaled scores, for each CLAST subtest for specified periods of time (table 2).

TABLE 2

Standards (Passing Scores) for CLAST Subtests

	Scaled Scores			
Time Period	Essay	English Language Skills	Reading	Mathematics
8/1/84 - 7/31/86	4	265	260	260
8/1/86 - 7/31/89	4	270	270	275
8/1/89 - 7/31/91	4	295	295	285
8/1/91 and thereafter	5	295	295	295

These tiers of standards are viewed by state-level panel members as reasonable expectations for all students, given the instructional program available to students taking the CLAST during each time period. The CLAST Technical Report, 1983-84 provides a full description of the process through which the standards were developed.

III. DESCRIPTION OF THE CLAST

Each form of the CLAST is developed according to specific guidelines which ensure that test forms from one administration to another are parallel in content and that administration procedures are standardized. This chapter describes the guidelines.

Test Specifications

For each of the three annual administrations (fall, winter, and spring), a different test is created; however, each test measures the same number of items in each broad skill area (table 3). To increase test security, two forms of each test are printed for each administration. Both forms contain the same scored items, but the order of item placement is different in each form. Developmental items are embedded in each test form in order to collect data needed to add items to the item bank.

The CLAST is comprised of four subtests. The Essay subtest is presented in a four-page folder; the English Language Skills and Reading subtests are in the same test book, and the Mathematics subtest is in separate test book.

Item Bank

As items are developed, they are numbered with a nine-digit code identifying the subtest, skill, sequence number, and graphic. These items are stored in a card file and a word processing file that are updated as items are revised. New items are added to the bank following the review of the developmental items from each administration.

A history and attribute computer file is kept for the item bank and used in the selection of items for test forms and in the test analysis process. The file includes attributes such as the item code, broad skill code, item flag, date used, and test form. Statistical data include the percentage correct, item point-biserial coefficient, Rasch difficulty, fit statistics, and index of discrimination for each item. Data on items are kept in the active file for six administrations. After that time, a hard copy and a tape record are stored. The computer bank then is rotated to remove the data from the earliest administrations.

Test Assembly

For each administration items are drawn from the item bank to meet the test specifications. Items are selected to minimize the difference in difficulty between forms. Current item difficulty values are used in the selection process. Test form item difficulties are centered near zero logits. Small variations in mean difficulty occur, particularly in the reading test where items are tied to specific passages. Alternate forms are adjusted to the common scale by the equating procedures described in Chapter IV.

TABLE 3

CLAST Specifications by Subtest, 1989-90

Subtest and Broad Skill	Number of Skills	Number of Items		
		Scored	Developmental	Total
ESSAY				
Two essay topics are provided; the examinee chooses one on which to write. Many general writing skills are tested including the eleven tested on the English Language Skills subtest.				
ENGLISH LANGUAGE SKILLS				
Word Choice	2	6		
Sentence Structure	4	13		
Grammar, Spelling, Capitalization, and Punctuation	5	16		
Total	11	35	5	40
READING				
Literal Comprehension	3	9		
Critical Comprehension	8	27		
Total	11	36	5	41
MATHEMATICS				
Arithmetic	13	13		
Algebra	16	16		
Geometry and Measurement	10	7		
Logical Reasoning	8	7		
Statistics, including Probability	9	7		
Total	56	50	5	55

The plan for format and arrangement of items in test forms is intended to make each form attractive and easy to read. Multiple-choice writing items are grouped by format and content to make test time efficient for students.

Test Instructions

General instructions provided to students contain information about scoring, recording answers, number of items, and time allotted for each subtest. Directions state that scores are based on the number of right answers with no correction for guessing.

The CLAST was administered in one session, which required nearly five hours. Although actual test time was four hours, additional time was required to check in examinees, code identifying information, distribute and collect materials, read directions for each subtest, and provide a ten-minute restroom break. The essay test was administered first, and students were allowed 60 minutes to complete it; the English language skills and reading tests were given next and 80 minutes were allowed for their completion; the computation test was administered last, and students were given 90 minutes to work on it.

Modifications in test format, such as braille, audio cassette and large print, were available for handicapped students. In addition, the test schedule and administration procedures were modified for handicapped examinees. Details of these modifications are provided in the CLAST Institutional Test Administrator's Manual.

Quality Control

Test form quality is maintained through an extensive review process. Drafts of new test forms are reviewed by staff of the Technical Support Contractor and the Department of Education. After changes in items and corrections are made, there is a thorough review of camera-ready copy, which is followed by a careful review of bluelines. Additional information about the performance of the test is taken from the institutional test administrators' and room supervisors' reports and on-site visits to test centers by Department of Education personnel. These reports provide information about the quality of test booklets, the standardization of test administrations, and the adequacy of allotted test times.

IV. TECHNICAL CHARACTERISTICS OF THE CLAST

To preserve comparability of CLAST scores from one administration to the next, test scores are equated using a base scale. To ensure reliability and validity of the test and test items, many traditional test analysis procedures are used. This section describes the equating process and procedures used to review the reliability and validity of the test.

Test Score Equating

The Rasch Model

The CLAST scale development is based on the logistic response model of Georg Rasch, presented in Probabilistic Models for Some Intelligence and Attainment Tests, 1960. Rasch describes a probabilistic model in which the probability that a person will answer an item correctly is assumed to be based on the ability of a person and the difficulty of the item. These estimates are derived independently and are not related to the particular sample of people or of items. When the assumptions of the model are met, tests of unequal difficulty can be equated.

Rasch model estimates of person ability and item difficulty are obtained using the unconditional maximum likelihood estimation procedure described in Wright, Mead, and Bell, BICAL: Calibrating Items With the Rasch Model, 1980. The probability of a score X_{vi} is expressed as

$$P(X_{vi} \mid B_v, \delta_i) = \frac{\exp [X_{vi}(B_v - \delta_i)]}{1 + \exp [B_v - \delta_i]}$$

where X_{vi} = a score, B_v = person ability, and δ_i = item difficulty.

Person ability in logits represents the natural log odds for succeeding on items which define the scale origin. The item difficulty in logits represents the natural log odds for failure on an item by persons with abilities at the scale origin.

One key assumption of the Rasch model is that a test under consideration is unidimensional. That is, it measures only one underlying student cognitive ability. Unfortunately, ability is considered to be "latent" and cannot be seen or measured in a very precise manner. Therefore, it is important to monitor the performance of the test and to conduct studies which will indicate whether the test is likely to be unidimensional. This has been done with the CLAST examination in two studies. The first study was performed in 1984 with the computation test. The second was done in 1986 with the reading, computation, and writing tests. Both studies showed that the use of Rasch techniques is justified.

Calibration of Items

Item difficulties are obtained by calibrating the scored items for each administration. Three systematic random samples of 700 records are drawn. The items are calibrated, and the item difficulty logits are averaged from

the three calibration samples. Using the averaged difficulties, the item logits are adjusted to the October 1982 base scale.

Item history records are kept in a computer file and updated after each administration. The stability of Rasch difficulty, discrimination values, and fit statistics are checked, and items that change values by more than .3 logit are flagged for further inspection. In addition, following each administration, items are re-examined against established item screening criteria.

Newly developed or revised items are embedded within each form of the test and then calibrated and adjusted to the base scale. These items are not counted toward examinees' scores and are not included in the initial calibrations used to develop the score scale. After the score scale is created, each test form is recalibrated with both the new and the scored items to estimate item difficulties of the new items. The scored items serve as a link between the new items in each test form. Item difficulties for the new items are adjusted to the base scale using the linking constant derived from the comparison of the calibration of the scored items to their base item difficulties. For a complete discussion of the method, see Ryan, J., Equating New Test Forms to an Existing Test, 1981.

Generation of Ability Estimates

The traditional estimate of achievement level is the raw score obtained from the number of correct answers provided. The Rasch model is used to generate ability estimates corresponding to the traditional test score.

Adjusted item difficulty logits obtained in item calibration become the basis for estimating person abilities. Generation of ability estimates results in a logit ability scale corresponding to the logit difficulty scale of items. Rasch ability logits are derived using the unconditional maximum likelihood estimation procedures of the program ABIL-EST (Ryan, 1981).

The ability estimate corresponding to each raw score between one point and the number of items minus one is calculated. (Perfect or zero scores are not included in Rasch calculations.) The ability logit scale is then centered at the mean for the October 1982 administration and converted to the standard score scale using a linear transformation.

Linking Scaled Scores

Through use of Rasch methodology, it is possible to place scores from tests of unequal difficulty on the same scale. While the CLAST difficulty is controlled by selecting items having approximately the same average and range of difficulty for each administration, some fluctuation in difficulty may occur in order to use items representing a broad range of content and difficulty. Differences in test form difficulty are controlled by equating.

Tests forms given on two different occasions are equated by using information obtained from a subset of items common to both forms. These common items are known as "anchor items." The performance of the two groups of examinees on the anchor items is used to adjust the measurement scales for

the two forms; the measurement scale for the second form is "adjusted" to that of the first form. From a measurement perspective, the examinees in both instances took the same form of the test. For the CLAST, all test forms are equated back to the first administration of October 1982. With this approach, all students face identical hurdles in that no student has the advantage of an "easier" form.

For each administration, CLAST item difficulties have been adjusted to the base scale of October 1982. Item logits obtained from calibrating the scored items are adjusted by adding the linking constant to each item logit. The difference in average difficulty represents the shift in overall difficulty between test forms. This constant is added to the current item logits to adjust them to the base scale. The stability of the link is evaluated by comparing the difficulty values over time to the values in the base scale.

Reliability of Scores

Reliability is an indicator of the consistency in measurement of student achievement. It provides an estimate of the variation in results that can be attributed to random error in measurement. The index of reliability is interpreted as the ratio of true-score variance to observed-score variance. Reliability is estimated somewhat differently for multiple-choice scores and essay ratings. Procedures used with each type of score are described in the following sections.

Reliability of Multiple-Choice Scores

The reliability of multiple-choice subtest scores is estimated using the Kuder-Richardson Formula 20 (KR-20) coefficient and the standard error of measurement (SEM). The KR-20 coefficient is an internal consistency estimate of reliability, proposed by Kuder and Richardson in 1937, based on the concept that achievement on items drawn from the same content domain should be related. The formula reported as the KR-20 is

$$rtt = \frac{k}{k-1}\left[\frac{s_t^2 - \Sigma\, pq}{s_t^2}\right]$$

where r_{tt} = estimated test reliability, k = number of test items, s_t^2 = variance of examinees' total scores, and $\Sigma\, pq$ = sum of item variances.

The KR-20 coefficient is appropriate for estimating reliability of scores on multiple-choice tests. However, the KR-20 coefficient can be affected by the distribution of scores. For this reason, the SEM is also reported as an indicator of reliability for each multiple-choice subtest.

The SEM represents the expected standard deviation of scores for an individual taking a large number of randomly selected parallel tests. The mean of the set of scores would represent the individual's true score. Therefore, the SEM can be used to estimate confidence intervals around an individual's true score. Confidence intervals applied to obtained scores are not symmetrical about the obtained score, but the estimated true score is useful in obtaining the center for a confidence zone to be used with

the obtained score. The smaller the SEM, the less dispersed are the parallel test scores and the more likely the estimate is close to the individual's true score.

The formula for computing the SEM is $SEM = s_t\sqrt{1-r_{tt}}$ where s_t = standard deviation of the test scores and r_{tt} = test reliability coefficient.

The KR-20s and SEMs for the CLAST multiple-choice subtests indicate they are acceptably reliable (table 4).

TABLE 4

Multiple-Choice Raw Score Reliability Statistics, 1989-90

	English Language Skills			Reading			Mathematics		
	October	March	June	October	March	June	October	March	June
KR-20	.70	.73	.73	.71	.81	.81	.86	.88	.86
SEM	1.88	1.76	2.02	2.14	2.23	2.34	2.82	2.84	2.85

Reliability of Essay Ratings

Reliability of essay ratings is evaluated in several ways to ensure that raters have adhered to established criteria for scoring essays. Consistency in scoring is maintained by training the raters and monitoring the scoring process; the reliability of the combined ratings is estimated by coefficient alpha. Both procedures are described below.

Training prior to and during scoring is used to develop and maintain consistency in scoring of the individual rater and the group of raters. The scoring process is monitored by checking the assignment of ratings, the number of split ratings, and the distribution of ratings of each reader. All papers assigned non-contiguous ratings are submitted to a referee who resolves the split scores. During and after each reading session, reader agreement data reflecting the reliability of ratings are reviewed. For the 1989-90 test administrations, the percentage of complete agreement between readers for all papers ranged from 55.8 to 57.6, while the percentage of non-contiguous scores ranged from 2.0 to 2.4 (table 5). These data show that over 97 percent of all the ratings were identical or contiguous (within one point of each other), indicating a high level of reader agreement. The complete agreement, by topic, resulting from the assignment to a referee of papers with non-contiguous scores was between 64 and 68% (table 6).

TABLE 5

Summary Data for All Essay Readers, 1989-90

	October		March		June	
	Number	Percent	Number	Percent	Number	Percent
Total Papers Read	18,630	100.0	27,412	100.0	12,449	100.0
Non-Contiguous Scores	446	2.4	624	2.3	245	2.0
Total Agreement Between Readers	10,392	55.8	15,549	56.7	7,165	57.6

TABLE 6

Essay Reader Agreement after Referee, 1989-90

	% Complete Agreement			% Agreement within One Point		
	October	March	June	October	March	June
Topic 1	62	64	68	38	36	32
Topic 2	64	64	64	36	36	36

Reliability of combined ratings for essays is estimated by coefficient alpha, which gives the expected correlation between combined ratings of the scoring team and those of a hypothetical parallel team doing the same task. The formula is

$$r_{kk} = \frac{k}{k-1}\left[1 - \frac{\Sigma s_i^2}{s_t^2}\right]$$

where r_{kk} = coefficient of reliability, k = number of test items, Σs_i^2 = sum of item variances, and s_t^2 = variance of examinees' total scores.

Alpha coefficients by topic for the ratings from 1989-90 show they are consistent across topics and administrations (table 7).

TABLE 7

Alpha Coefficients, 1989-90

	Non-Refereed Scores			Refereed Scores		
	October	March	June	October	March	June
Topic 1	.75	.74	.72	.85	.83	.82
Topic 2	.70	.74	.73	.80	.83	.82

Reliability of Pass/Fail Classification

Since CLAST scores are used to determine whether students in Florida's community colleges and universities have achieved the level of performance required for the award of an associate in arts degree or for admission to upper division status, reliability in testing and retesting is an important issue. The reliability issue of interest is whether students would consistently pass or would consistently fail if several parallel forms of the test were administered to them. The results of a test-retest study conducted in 1984 indicate that the CLAST is reliable for making pass/fail decisions based on the 1984-86 standards. A complete report of the study is available from the Department of Education and a summary is available in Appendix F.

Item Analysis

An item analysis as shown in figure 1 is prepared for the total group of examinees, each gender, and each racial/ethnic category. These analyses include number and percentage of examinees who chose each item response, omitted the item, or gridded more than one response. In addition, they include item difficulty (proportion of examinees choosing the correct response), item discrimination, and point biserial correlation.

Following test administration, preliminary item analyses are run on the first answer sheets received for scoring. Results of these analyses are screened for item flaws or key errors. Clues to such errors are low discrimination indices or Rasch fit statistics with high values. Other indicators of problems include lack of balance in foil distributions or inordinate difficulty. Items exhibiting these characteristics are flagged and, following a Department of Education review, may be excluded from scoring.

Pretesting new items embedded in the test forms is another form of quality control. Before an item is added to the bank, it is pretested as a non-scored item, and its item statistics are reviewed. Items not meeting the item selection criteria are examined to determine if they are adequate measures of the skills. Any item deemed inappropriate is flagged and not used on the CLAST.

COMPUTATION SCORE

ITEM NUMBER	ITEM RESPONSES A	B	C	D	E	OMIT	MULT	ITEM RESPONSE FIGURES ARE TOTALS, NOT PERCENTAGES ITEM DIFFICULTY	ITEM DISCRIMINATION	BISERIAL CORRELATION
1	936	1088	7270+	989	0	87	1	0.70	0.52	0.45
2	2679	458	5012+	2185	0	36	1	0.48	0.46	0.36
3	1175	1743	1211	6189+	0	51	2	0.60	0.66	0.52
4	7528+	1004	629	1165	0	45	0	0.73	0.43	0.39
5	2945+	1389	3305	2661	0	71	0	0.28	0.33	0.29
6	1835	5650+	957	1859	0	70	0	0.54	0.39	0.32
7	1614	733	7020+	963	0	41	0	0.68	0.39	0.34
8	724	1472	1423	6694+	0	57	1	0.65	0.58	0.48
9	70	80	124	10071+	0	24	2	0.97	0.07	0.21
10	78	4132	171	5961+	0	29	0	0.57	0.59	0.48
11	300	465	1785	7775+	0	40	6	0.75	0.36	0.34
12	1513	1823	4068+	2926	0	39	2	0.39	0.43	0.36
13	538	8493+	190	1127	0	23	0	0.82	0.42	0.46
14	737	487	6102+	3003	0	41	1	0.59	0.39	0.31
15	4602+	1530	2269	1896	0	74	0	0.44	0.44	0.35
16	628	445	8133+	1139	0	26	0	0.78	0.48	0.47
17	1607	5457+	1259	2011	0	37	0	0.53	0.49	0.39
18	1905	4376+	1188	2825	0	76	1	0.42	0.36	0.29
19	3908	6169+	174	73	0	47	0	0.59	0.51	0.41
20	497	2837	6299	674	0	64	0	0.61	0.58	0.47
21	9838+	137	144	219	0	33	0	0.95	0.13	0.26
22	253	2659	7380+	45	0	34	0	0.71	0.49	0.42
23	2282	256	383	7440+	0	9	1	0.72	0.35	0.32
24	977	4073+	3507	1779	0	34	1	0.39	0.39	0.32
25	584	740	233	8794+	0	20	0	0.85	0.28	0.33
26	6820+	2975	380	182	0	14	0	0.66	0.42	0.36
27	1787	2828	3687+	2024	0	44	1	0.36	0.60	0.49
28	198	251	517	9374+	0	30	1	0.90	0.24	0.35
29	1568	791	1397	6592+	0	23	0	0.64	0.58	0.48
30	816	3564	459	5514+	0	17	1	0.53	0.34	0.28
31	370	1509	7646+	784	0	62	0	0.74	0.37	0.35
32	7175+	796	2337	33	0	29	1	0.69	0.36	0.32
33	5775	671	3582+	325	0	18	0	0.35	0.61	0.50
34	518	627	1169	8031+	0	25	1	0.77	0.36	0.35
35	5130+	1463	2825	909	0	43	1	0.49	0.48	0.38
36	3530	350	5665+	808	0	16	2	0.55	0.39	0.32
37	880	2143	6791+	516	0	41	0	0.65	0.55	0.46
38	1224	448	1736	6944+	0	17	2	0.67	0.55	0.47
39	5442+	2579	1025	1265	0	58	2	0.52	0.45	0.37
40	573	2231	1834	5675+	0	58	0	0.55	0.51	0.41
41	7121+	1489	1132	535	0	93	1	0.69	0.60	0.52
42	3056	2606+	3938	697	0	74	0	0.25	0.39	0.36
43	917	1164	6223+	1980	0	86	1	0.60	0.61	0.49
44	879	2157	1366	5861+	0	107	1	0.57	0.65	0.53
45	569	1890	1322	6541+	0	49	0	0.63	0.62	0.51
46	1207	5065+	1266	2797	0	35	1	0.49	0.63	0.50
47	253	363	5018+	4708	0	29	0	0.48	0.33	0.27
48	1435	1810	2554	4531+	0	41	0	0.44	0.58	0.45
49	2102	4835+	1823	1426	0	182	3	0.47	0.28	0.23

\+ - INDICATES CORRECT ANSWER
** - INDICATES EVERYONE GIVEN CREDIT
* - INDICATES QUESTION THROWN OUT

Fig.1. Example of an item analysis

Preventing Item Bias

In addition to examining item analyses, review panels established at each stage of test development considered the issue of bias in the items. Scatter graphs were examined after each administration to determine if particular items operated differently for various racial or ethnic groups.

A scatter graph (fig. 2) contrasts performance on individual items by racial/ethnic or gender categories. An item difficulty is identified as an outlier if it deviates substantially from the general relationship for the compared groups. Consistent differences in item difficulties may indicate only a difference in the level of achievement for the compared groups, but items that deviate from this general pattern are further examined for content bias that may be related to gender or racial/ethnic background.

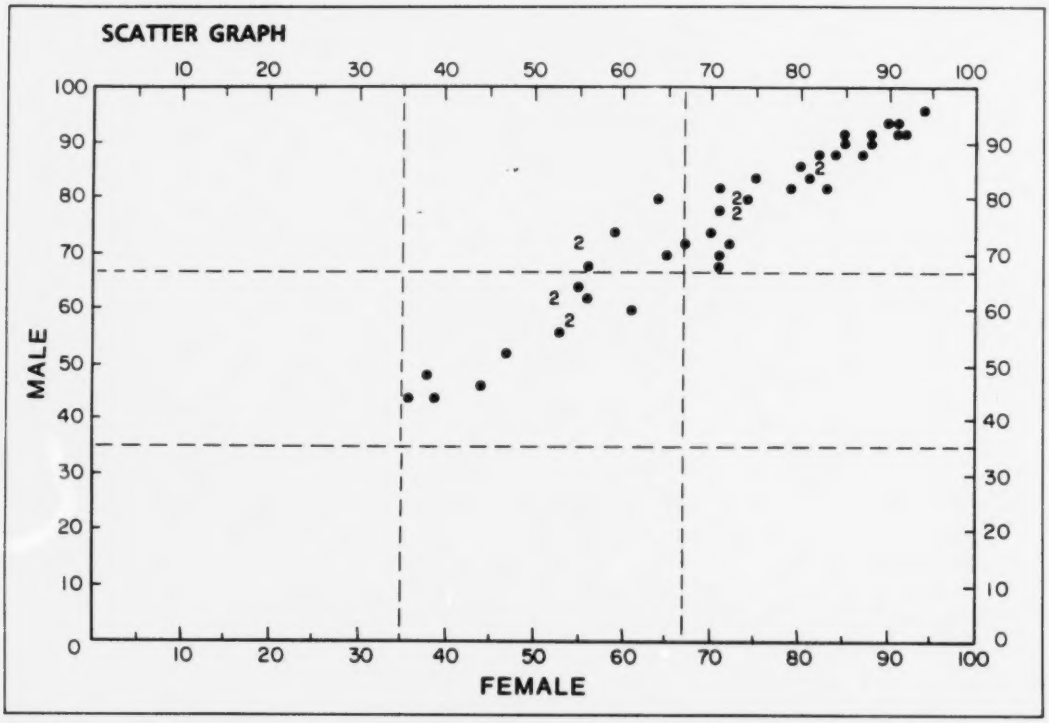


Fig. 2. Example of a scatter graph of item difficulties comparing the performance of males with that of females.

Validity of Scores

Strictly speaking, one should not describe a test as being "valid." Instead, one should describe a test score as being "valid" for a particular purpose. Hence, test development operations are designed to build evidence for a particular type of score interpretation which is defined in advance.

Standards for Educational and Psychological Testing (1985) describes three types of validity: content, construct, and criterion. Content validity is the only important type for the CLAST because test scores are only interpreted in terms of what they indicate about student achievement of designated performance objectives. The CLAST does not measure a designated psychological characteristic (e.g., spatial visualization), so construct validity is not relevant. Further, as has been stated, the CLAST was not designed to predict a student's future performance in school. Hence, the criterion-related (i.e., predictive) validity is not relevant. Content validity is substantiated by determining the extent to which the test items adequately measure the specific skills they are designed to measure; that is, the extent to which the content of the test matches the set of skills. The validity of the test is established by following the plan and procedures for developing and selecting items for each form of the CLAST.

The general plan used in developing the test is outlined below.

1. General test specifications, consistent with the purpose of the CLAST, are developed by faculty who have expertise in both testing and the content areas (English language skills, reading, and mathematics) with assistance of Department of Education staff.

2. Item specifications detailing both content and format of items which can be developed to measure each of the skills, are developed by faculty with expertise in both the content areas and testing, with assistance of Department of Education staff.

3. Test items are written by faculty according to the guidelines provided by the item specifications and are reviewed by faculty and Department of Education staff with careful attention given to content, measurement, and bias issues.

4. Test items are field-tested in community colleges and state universities.

5. Items are analyzed statistically and selected for use in the test only if they meet criteria established by Department of Education staff and testing consultants.

6. A test plan for selection of items is followed in developing alternate forms of the test.

7. Scaled scores equated to the reference scale are generated using the Rasch model.

To summarize, validity of the test as a measure of achievement of the skills is established by following the plan for developing and selecting items. Content and testing specialists judge the adequacy of the items for measuring the skills, and the plan for selecting items ensures that each form of the CLAST is representative of the domain of skills being tested. Scores on each of the subtests, then, can be interpreted to be valid indicators of students' achievement of the communication and mathematics skills measured by the CLAST.

V. SCORING AND REPORTING PROCEDURES

Procedures for scoring the CLAST are designed to provide quality control and score scale stability for a testing program that has complex scoring and reporting requirements. The process for scoring and reporting reflects concern for reliability and comparability of the scores and for appropriate use of the scores. This chapter addresses those concerns.

Scoring Activities

Editing Answer Sheets

Following each administration, as answer sheets are received from each institution, they are edited for errors. Answer sheets are read by an NCS Sentry 7018 scanner programmed to identify mismarked or miscoded sheets. Each identified answer sheet is hand-checked and corrected according to the scoring conventions.

Rating sheets from holistic scoring of essays are also machine-scored. Editing procedures for holistic scoring include a verification of the legitimacy of reader numbers and score codes. Papers with invalid scores or with ratings that differ by more than one point are returned to the referee to be corrected and/or reviewed.

Scoring Conventions

Within the parameters of number-right scoring, certain conventions are observed: for a response to be considered valid, it must be recorded in the answer folder; for a score to be generated on a subtest, at least one response must be marked in the appropriate section of the answer sheet; and omits and double grids are counted as incorrect. To receive credit for the essay test, students must write on one of the two topics provided, and they must write the essay in their answer folders.

Students' subtest scores below the chance level are compared to their other subtest scores. If a score is inconsistent with the student's performance on the other subtests, it is hand-checked to determine if the student entered the correct form code on the answer sheet.

Score Scales

A three-digit standard scaled score is generated for each administration for each of the multiple-choice subtests. The standard score scale is a linear transformation of the Rasch ability logits adjusted for the mean of the October 1982 administration. The formula used is

$$S_i = 30(X_a - C) + 300$$

where: S_i = scaled score, X_a = ability logit, C = October 1982 scale adjustment factor (1.87 for English language skills, 1.2 for reading and 1.0 for mathematics). Raw score to scaled score transformation data are generated for each subtest for each administration (tables 8, 9 and 10).

TABLE 8

English Language Skills Score Conversions, 1989-90

Raw Score	October		March		June	
	Ability	Scaled Score	Ability	Scaled Score	Ability	Scaled Score
0	-6.661	100	-6.811	095	-6.55	103
1	-5.698	129	-5.852	124	-5.59	132
2	-4.963	151	-5.119	146	-4.86	154
3	-4.507	164	-4.666	160	-4.40	168
4	-4.165	175	-4.326	170	-4.05	178
5	-3.886	183	-4.048	178	-3.77	186
6	-3.645	190	-3.809	185	-3.53	194
7	-3.432	197	-3.596	192	-3.31	200
8	-3.238	202	-3.403	197	-3.11	206
9	-3.058	208	-3.224	203	-2.93	212
10	-2.889	213	-3.056	208	-2.75	217
11	-2.728	218	-2.896	213	-2.59	222
12	-2.575	222	-2.744	217	-2.43	227
13	-2.426	227	-2.597	222	-2.27	231
14	-2.281	231	-2.454	226	-2.12	236
15	-2.140	235	-2.314	230	-1.98	240
16	-2.000	240	-2.177	234	-1.83	245
17	-1.862	244	-2.041	238	-1.69	249
18	-1.724	248	-1.906	242	-1.54	253
19	-1.586	252	-1.771	246	-1.40	258
20	-1.447	256	-1.635	250	-1.25	262
21	-1.306	260	-1.498	255	-1.11	266
22	-1.162	265	-1.359	259	-0.96	271
23	-1.014	269	-1.216	263	-0.80	276
24	-0.862	274	-1.068	267	-0.64	280
25	-0.703	278	-0.915	272	-0.47	285
26	-0.536	283	-0.754	277	-0.30	291
27	-0.358	289	-0.582	282	-0.11	296
28	-0.166	295	-0.397	288	0.09	302
29	0.045	301	-0.194	294	0.31	309
30	0.282	308	0.036	301	0.56	316
31	0.557	316	0.303	309	0.85	325
32	0.895	326	0.632	318	1.21	336
33	1.346	340	1.075	332	1.67	350
34	2.076	362	1.795	353	2.42	372
35	3.031	390	2.736	382	3.40	402

TABLE 9

Reading Score Conversions, 1989-90

Raw Score	October		March		June	
	Ability	Scaled Score	Ability	Scaled Score	Ability	Scaled Score
0	-6.169	114	-5.653	130	-5.75	127
1	-5.201	143	-4.711	158	-4.79	156
2	-4.463	166	-3.990	180	-4.06	178
3	-4.003	179	-3.547	193	-3.60	192
4	-3.658	190	-3.218	203	-3.27	201
5	-3.375	198	-2.951	211	-2.99	210
6	-3.131	206	-2.723	218	-2.75	217
7	-2.914	212	-2.521	224	-2.55	223
8	-2.716	218	-2.339	229	-2.36	229
9	-2.532	224	-2.171	234	-2.18	234
10	-2.358	229	-2.014	239	-2.02	239
11	-2.194	234	-1.865	244	-1.86	244
12	-2.035	238	-1.723	248	-1.72	248
13	-1.882	243	-1.586	252	-1.57	252
14	-1.733	248	-1.453	256	-1.43	257
15	-1.587	252	-1.324	260	-1.30	261
16	-1.443	256	-1.196	264	-1.16	265
17	-1.300	261	-1.070	267	-1.03	269
18	-1.158	265	-0.945	271	-0.90	273
19	-1.016	269	-0.820	275	-0.77	276
20	-0.874	273	-0.695	279	-0.64	280
21	-0.730	278	-0.568	282	-0.51	284
22	-0.584	282	-0.439	286	-0.37	288
23	-0.435	286	-0.307	290	-0.23	293
24	-0.282	291	-0.171	294	-0.09	297
25	-0.123	296	-0.030	299	0.06	301
26	0.042	301	0.117	303	0.22	306
27	0.215	306	0.272	308	0.38	311
28	0.400	312	0.438	313	0.56	316
29	0.599	317	0.618	318	0.75	322
30	0.818	324	0.817	324	0.97	329
31	1.065	331	1.041	331	1.21	336
32	1.351	340	1.305	339	1.50	345
33	1.700	351	1.630	348	1.85	355
34	2.165	364	2.068	362	2.32	369
35	2.912	387	2.785	383	3.08	392
36	3.891	416	3.721	411	4.07	422

TABLE 10

Mathematics Score Conversions, 1989-90

Raw Score	October		March		June	
	Ability	Scaled Score	Ability	Scaled Score	Ability	Scaled Score
0	-5.948	121	-6.044	118	-6.17	114
1	-5.023	149	-5.113	146	-5.24	142
2	-4.313	170	-4.398	168	-4.52	164
3	-3.883	183	-3.966	181	-4.09	177
4	-3.569	192	-3.649	190	-3.77	186
5	-3.318	200	-3.395	198	-3.51	194
6	-3.106	206	-3.182	204	-3.29	201
7	-2.921	212	-2.995	210	-3.10	207
8	-2.757	217	-2.829	215	-2.93	212
9	-2.607	221	-2.677	219	-2.78	216
10	-2.468	225	-2.538	223	-2.63	221
11	-2.339	229	-2.407	227	-2.50	225
12	-2.218	233	-2.285	231	-2.37	228
13	-2.102	236	-2.169	234	-2.25	232
14	-1.992	240	-2.058	238	-2.14	235
15	-1.886	243	-1.951	241	-2.03	239
16	-1.784	246	-1.848	244	-1.92	242
17	-1.684	249	-1.748	247	-1.82	245
18	-1.587	252	-1.650	250	-1.71	248
19	-1.492	255	-1.555	253	-1.62	251
20	-1.399	258	-1.461	256	-1.52	254
21	-1.307	260	-1.369	258	-1.42	257
22	-1.216	263	-1.278	261	-1.33	260
23	-1.126	266	-1.187	264	-1.24	262
24	-1.036	268	-1.097	267	-1.14	265
25	-0.946	271	-1.007	269	-1.05	268
26	-0.856	274	-0.917	272	-0.96	271
27	-0.766	277	-0.827	275	-0.86	274
28	-0.675	279	-0.736	277	-0.77	276
29	-0.583	282	-0.644	280	-0.68	279
30	-0.489	285	-0.551	283	-0.58	282
31	-0.394	288	-0.456	286	-0.49	285
32	-0.297	291	-0.359	289	-0.39	288
33	-0.198	294	-0.260	292	-0.29	291
34	-0.096	297	-0.159	295	-0.19	294
35	0.010	300	-0.054	298	-0.08	297
36	0.120	303	0.055	301	0.02	300
37	0.234	307	0.169	305	0.14	304
38	0.355	310	0.288	308	0.25	307
39	0.483	314	0.413	312	0.38	311
40	0.619	318	0.547	316	0.51	315
41	0.765	322	0.690	320	0.65	319
42	0.925	327	0.845	325	0.80	324
43	1.101	333	1.016	330	0.97	329
44	1.299	338	1.207	336	1.15	334
45	1.527	345	1.425	342	1.37	341
46	1.798	353	1.684	350	1.62	348
47	2.135	364	2.006	360	1.94	358
48	2.592	377	2.444	373	2.37	371
49	3.335	400	3.165	394	3.08	392
50	4.306	429	4.105	423	4.00	420

The score scale ranges from approximately 100 points to 400 points. It is centered at 300 points designating the state average score on the October 1982 administration. All subsequent examinations are equated to this administration. Differences in scaled score ranges across test forms occur as a result of differences in the range of item difficulty in test forms. The difficulty of each form is controlled, however, so that these shifts in the average score range are small. If one test form has items that are more difficult, it is possible to obtain a higher scaled score because the harder items measure a higher level of achievement.

The essay score is assigned on a scale of two to eight points. Two readers rate each essay on a rating scale from one to four points. The essay score is the sum of the two ratings. The holistic scoring procedure and rating scale are discussed in the next section.

Essay Scoring

Holistic scoring or evaluation, a process for judging the quality of writing samples, has been used for many years by testing agencies in credit-by-examination, state assessment, and teacher certification programs.

Holistic Scores

Essays are scored holistically—that is, for the total, overall impression they make on the reader—rather than analytically, which requires careful analysis of specific features of a piece of writing. Holistic scoring assumes that the skills which make up the ability to write are closely interrelated and that one skill cannot be separated from the others. Thus, the writing is viewed as a total work in which the whole is something more than the sum of the parts. A reader reads a writing sample once, forms an impression of its overall quality, and assigns it a numerical rating based on his/her judgment of how well the paper meets a particular set of established criteria. A four-point scale reflecting the following performance levels is used to score CLAST essays.

Score of 4: Writer purposefully and effectively develops a thesis. Writer uses relevant details, including concrete examples, that clearly support generalizations. Paragraphs carefully follow an organizational plan and are fully developed and tightly controlled. A wide variety of sentences occurs, indicating that the writer has facility in the use of language, and diction is distinctive. Appropriate transitional words and phrases or other techniques make the essay coherent. Few errors in syntax, mechanics, and usage occur.

Score of 3: Writer develops a thesis but may occasionally lose sight of purpose. Writer uses some relevant and specific details that adequately support generalizations. Paragraphs generally follow an organizational plan and are usually unified and developed. Sentences are often varied, and diction is usually appropriate. Some transitions are used, and parts are usually related to each other in an orderly manner.

Syntactical, mechanical, and usage errors may occur but usually do not affect clarity.

Score of 2: Writer may state a thesis, but the essay shows little, if any, sense of purpose. Writer uses a limited number of details, but they often do not support generalizations. Paragraphs may relate to the thesis but often will be vague, underdeveloped, or both. Sentences lack variety and are often illogical, poorly constructed, or both. Diction is pedestrian. Transitions are used infrequently, mechanically, and erratically. Numerous errors may occur in syntax, mechanics, and usage and frequently distract from clarity.

Score of 1: Writer's thesis and organization are seldom apparent, but, if present, they are unclear, weak, or both. Writer uses generalizations for support, and details, when included, are usually ineffective. Undeveloped, ineffective paragraphs do not support the thesis. Sentences are usually illogical, poorly constructed, or both. They usually consist of a series of subjects and verbs with an occasional complement. Diction is simplistic and frequently not idiomatic. Transitions and coherence devices, when discernible, are usually inappropriate. Syntactical, mechanical, and usage errors abound and impede communication.

Holistic Scoring

The holistic scoring session must be conducted in a highly organized manner with competent staff members who have clearly specified responsibilities. For ten thousand essays, the holistic scoring staff consists of a chief reader, three assistant chief readers, twenty table leaders, and one hundred readers. A support staff of a manager and five clerks is also required.

The scoring procedure follows this pattern. Prior to the scoring session, the chief reader and assistants sample the total group of essays to choose from each of the two topics examples which clearly represent the established standards for each of the four ratings on the rating scale. These essays are known as range finders. In addition, other essays are chosen as training materials during the scoring sessions.

After range finders and samples are selected, table leaders meet with the chief and assistant chief readers to score the samples and determine if the samples clearly represent the four levels of the scale. The purpose of this session is to refine the sample selection and to ensure consensus among table leaders. Range finders from previous administrations are also reviewed and used in the training to ensure consistency in scoring from one administration to another.

Immediately prior to and intermittently throughout the scoring session, the chief reader trains the readers using the range finders and other samples. Immediately after the initial training session, scoring begins. Each essay is read by two readers who assign it a rating of one, two,

three, or four. The sum of the ratings is the total score assigned to the essay. A total of four or above is passing.

When the total score is three, the essay is read by a third reader called a referee. The referee's rating will match one of the other ratings and replace the nonmatching one. The new total score is either four, which is passing, or two, which is not passing; a total score of three is not reported. When the ratings of two readers of the same essay are not contiguous, the essay is also refereed.

A more complete description of the process is in Procedures for Conducting Holistic Scoring for the Essay Portion of the College-Level Academic Skills Test available in the Department of Education office.

Recruitment of Readers

Each institution that registers students for the CLAST may participate in the holistic scoring process. The chief reader solicits nominations for readers from the chairs of English departments in community colleges and universities. Nominations for readers are made on the basis of the candidate's interest in the process, willingness to set aside personal standards for judging the quality of writing and to undergo training, and availability to work over weekends. Candidates must have a minimum of two years' experience teaching composition, hold at least a master's degree or equivalent, have a major in English in at least one baccalaureate degree, and teach composition as part of their assigned responsibilities. Nominations may include secondary school teachers who teach composition at the junior or senior year level in high schools and faculty who teach composition in private postsecondary institutions.

Upon receiving nominations from department chairs, the chief reader and the Technical Support Contractor ask each nominee interested in becoming a reader to complete and submit an application form. The forms are used to determine whether applicants meet the criteria for readers.

Reporting Test Results

The reports outlined below are generated for each administration. In addition to these reports, institutions may request from the Technical Support Contractor a computer tape or diskettes containing their students' data, including item responses. Thus, institutions can generate their own reports and update files of students' records. A test blueprint giving item-skill correspondence and a data tape format are also provided to institutions.

Student Reports

The individual student report (fig. 3) and a score interpretation guide are mailed to students approximately six weeks after the examination date. A scaled score is reported for each subtest taken. In the boxes to the right of the scale score is reported the percentage of items correct in each broad skill area. Although the percentages are reported to the student, they do not become part of the student's transcript. The percent-

Individual Score Report
COLLEGE-LEVEL ACADEMIC SKILLS TEST
DATE OF EXAM:

S.S. #

INSTITUTION

Following are your results for the College-Level Academic Skills Test. The enclosed interpretation guide will help you understand your scores. The score in the first box below is your essay grade. The three-digit numbers listed first in the three remaining boxes are your scale scores for each subtest. After each scale score you will find the percent of items you answered correctly for each of the broad skill areas within the subtest. This report is provided for your information. The official record of your scores will be kept by your institution on your transcript.

ESSAY

Essay Rating

ENGLISH LANGUAGE SKILLS

SCALE SCORE	Word Choice	Sentence Structure	Grammar, Spelling, Punctuation, Capitalization

READING

SCALE SCORE	Comprehension	
	Literal	Critical

MATHEMATICS

SCALE SCORE	Arithmetic	Algebra	Geometry - Measurement	Logical Reasoning	Statistics

Passing scores on CLAST have been established by the State Board of Education as follows:

	Essay	English Language Skills	Reading	Mathematics
8/1/84 - 7/31/86	4	265	260	260
8/1/86 - 7/31/89	4	270	270	275
8/1/89 - 7/31/90	4	295	295	285

Students are required to meet the standards in effect at the time they first took the test.

If you have questions about your scores, you should contact:

Fig. 3. Copy of a blank student report form.

BEST COPY AVAILABLE

ages help students determine their relative strengths and weaknesses in the broad skill areas represented on the test.

Preliminary Reports—prepared at the state and institutional levels

1. Summary statistics (means, medians, and standard deviations) and frequency distributions of scores by

   a. Student classification
      - Community college A.A. program
      - Community college A.S. program
      - University native student
      - University transfer student

   b. Racial/ethnic classification:
      - White/non-Hispanic
      - Black/non-Hispanic
      - Hispanic
      - American Indian/Alaskan native
      - Asian/Pacific Islander
      - Non-Resident Alien

   c. Gender by racial/ethnic classification

2. Alphabetic roster of examinees' scores

Final Reports—prepared at the state and institutional levels

1. Means and percents of first-time examinees meeting current standards for

   a. students with 60 or more hours
   b. students with fewer than 60 hours
   c. state university native students
   d state university transfer students
   e. students by gender and racial ethnic category for each institution, all public institutions, all private institutions, all community colleges, and all state universities

2. Means and percents of first-time examinees meeting future standards by gender and racial ethnic category for each institution, all public institutions, and all private institutions

3. Means and percents of retake examinees meeting required standards by gender and racial ethnic category for each institution, all public institutions, and all private institutions

Statistical Reports—prepared at the state level only

1. Rasch item calibrations and fit statistics
2. Scaled score derivations
3. Classical item analysis by racial/ethnic classification
4. Item difficulty plots by gender and racial/ethnic classification
5. KR-20 coefficients and SEM's for multiple-choice subtests
6. Interrater reliability for essay scores

7. Coefficient alpha by gender and racial/ethnic classification for essay scores

Interpreting and Using Scores

CLAST scores are reported to indicate students' achievement of those skills upon which the test is based. The CLAST scaled scores, not the raw scores, for each subtest are used for this purpose since the scaled scores have been adjusted for differences in difficulty in test forms. A scaled score of 300, for instance, represents the same achievement level across forms but may require a higher raw score on an easier form than on a harder one. The same scaled score, then, represents the same level of achievement of the skills regardless of the test form taken.

The use of CLAST scores is prescribed by Florida Statutes and Rules of the SBE. Use of scores prior to August 1, 1984, was limited to student advising and curriculum improvement. Since August 1, 1984, students in public institutions in Florida are required to have CLAST scores which satisfy the standards set forth in Rule 6A-10.0312, FAC, for the award of an associate in arts degree and for the admission to upper division status in a state university in Florida. However, students who have satisfied CLAST standards on three of the four subtests and who are otherwise eligible may be enrolled in state universities for up to an additional thirty-six semester credits of upper division course work before they are required to pass the fourth subtest.

Standards (passing scores) for the CLAST have been adopted by the SBE in Rule 6A-10.0312(1), FAC. The standards for each designated period of time are indicated in Chapter II.

The CLAST was not developed to predict success in upper division programs, but to assess the level of achievement of the skills listed in Appendix A. Any use of the scores for selection of students for specific upper division programs must be empirically validated.

VI. SUMMARY OF 1989-90 RESULTS

The results of CLAST administrations indicate the level of achievement of communication and computation skills by students in community colleges and state universities. Summary data presented in this section describe student performance on the CLAST as a whole and on each subtest. Summary data are based on only those students who were first-time takers in public institutions.

The mean, standard deviation, and median of raw scores and scaled scores are reported by subtest for each administration (table 11). Mean and median scaled scores for the June 1990 administration were consistently lower than their counterparts from either of the other two 1989-90 administrations; the mean and median raw scores, however, showed no consistent pattern.

Examinees who passed the CLAST are those who met the 1989 standards for each subtest. The percentage of examinees that passed the CLAST was 66 in October 1989 and was 59 in March and 56 in June of 1990 (table 12). The passing rates for groups of students classified on the basis of gender or racial/ethnic background varied across all administrations, ranging from a low of 24% in June to a high of 75% in October (table 12).

Mean scores are reported for all students, for students grouped according to gender, for students grouped according to racial/ethnic background, for students in community colleges, and for students in the state university system. These means are provided separately for the essay, English language skills, reading, and mathematics subtests and are found in tables 13, 14, 15, and 16, respectively.

TABLE 11

Raw and Scaled Scores, 1989-90
(First-Time Examinees in Public Institutions)

	No. of Items	Raw Score			Scaled Score		
		Mean	Std. Dev.	Median	Mean	Std. Dev.	Median
Essay							
October					4.9	1.4	5
March					4.9	1.4	5
June					4.8	1.4	5
English Language Skills							
October	35	30.2	3.4	31	317.6	30.9	316
March	35	31.0	3.4	32	318.8	32.2	318
June	35	28.8	3.9	29	314.7	30.7	309
Reading							
October	36	28.7	4.0	29	320.2	26.1	318
March	36	28.4	5.1	29	321.4	30.8	318
June	36	26.4	5.3	27	313.2	29.8	311
Mathematics							
October	50	36.0	7.5	37	307.7	28.4	306
March	50	35.7	8.2	37	305.2	31.0	305
June	50	35.9	7.6	37	304.2	28.6	304

TABLE 12

Percentage of Examinees Passing All Four Subtests, 1989-90
(First-Time Examinees in Public Institutions)

Examinee Group	October		March		June	
	Number Tested	Percent Passing	Number Tested	Percent Passing	Number Tested	Percent Passing
All	18,668	66	31,086	59	12,456	56
Male	7,993	66	13,224	59	4,903	57
Female	10,675	66	17,862	59	7,553	56
White	13,240	75	21,667	70	8,827	66
Black	2,060	42	3,381	31	1,126	24
Hispanic	2,320	42	4,257	35	1,812	35
Asian/Pacific Islander	496	49	848	47	307	37
American Indian/ Alaskan Native	36	75	79	47	31	55
Non-Resident Alien	389	41	655	32	263	31
Unknown Race	127	54	199	39	90	40
Community College	11,002	59	19,792	49	9,990	52
State University System	7,666	76	11,294	77	2,466	73

TABLE 13

Essay Mean Scaled Scores, 1989-90
(First-Time Examinees in Public Institutions)

Examinee Group	October		March		June	
	Number	Mean	Number	Mean	Number	Mean
All	18,723	4.9	31,180	4.9	12,495	4.8
Male	8,025	4.7	13,262	4.8	4,922	4.6
Female	10,698	5.0	17,918	5.1	7,573	4.9
White	13,276	5.2	21,724	5.2	8,848	5.1
Black	2,064	4.3	3,399	4.3	1,130	4.1
Hispanic	2,327	4.3	4,273	4.2	1,824	4.2
Asian/Pacific Islander	499	4.1	849	4.3	309	4.1
American Indian/ Alaskan Native	37	5.1	79	4.9	31	4.8
Non-Resident Alien	393	3.9	657	4.0	263	4.0
Unknown Race	127	4.6	199	4.4	90	4.3
Community College	11,032	4.7	19,854	4.7	10,021	4.7
State University System	7,691	5.2	11,326	5.4	2,474	5.2

TABLE 14

English Language Skills Mean Scaled Scores, 1989-90
(First-Time Examinees in Public Institutions)

Examinee Group	October		March		June	
	Number	Mean	Number	Mean	Number	Mean
All	18,752	318	31,207	319	12,520	315
Male	8,039	314	13,282	315	4,931	310
Female	10,713	321	17,925	321	7,589	317
White	13,293	323	21,737	325	8,851	321
Black	2,068	303	3,405	304	1,137	296
Hispanic	2,336	303	4,281	303	1,834	299
Asian/Pacific Islander	499	308	848	310	310	308
American Indian/ Alaskan Native	36	319	79	311	31	310
Non-Resident Alien	393	299	657	302	265	302
Unknown Race	127	310	200	303	92	302
Community College	11,047	313	19,878	313	10,043	312
State University System	7,705	325	11,329	329	2,477	324

TABLE 15

Reading Mean Scaled Scores, 1989-90
(First-Time Examinees in Public Institutions)

Examinee Group	October		March		June	
	Number	Mean	Number	Mean	Number	Mean
All	18,754	320	31,209	321	12,518	313
Male	8,040	321	13,281	322	4,931	315
Female	10,714	319	17,928	321	7,587	312
White	13,293	325	21,737	329	8,849	320
Black	2,069	304	3,406	302	1,139	292
Hispanic	2,336	308	4,281	305	1,833	299
Asian/Pacific Islander	500	309	848	311	309	298
American Indian/ Alaskan Native	36	328	79	320	31	311
Non-Resident Alien	393	302	657	301	265	299
Unknown Race	127	315	201	307	92	301
Community College	11,050	316	19,879	315	10,040	311
State University System	7,704	326	11,330	332	2,478	323

TABLE 16

Mathematics Mean Scaled Scores, 1989-90
(First-Time Examinees in Public Institutions)

Examinee Group	October		March		June	
	Number	Mean	Number	Mean	Number	Mean
All	18,750	308	31,163	305	12,520	304
Male	8,031	312	13,265	311	4,929	311
Female	10,719	304	17,898	301	7,591	300
White	13,288	312	21,707	312	8,854	309
Black	2,071	293	3,391	285	1,143	283
Hispanic	2,337	295	4,279	288	1,826	294
Asian/Pacific Islander	499	313	848	311	309	311
American Indian/ Alaskan Native	36	314	79	299	31	299
Non-Resident Alien	391	304	658	300	265	304
Unknown Race	128	302	201	291	92	296
Community College	11,047	303	19,842	298	10,038	302
State University System	7,703	315	11;321	317	2,482	314

BIBLIOGRAPHY

American Psychological Association. Standards for Educational and Psychological Testing. American Psychological Association: Washington, DC, 1985.

Brennan, R. L. and Kane, M. T. An index of dependability for mastery tests. Journal of Educational Measurement. 1977, 14, 277-298.

Florida Department of Education. CLAST Test Administration Manual for Institutional Test Administrators, 1986-87.

Florida Department of Education. CLAST Test Administration Plan, 1986-87.

Florida Department of Education. CLAST Technical Report, 1982-83.

Florida Department of Education. CLAST Technical Report, 1983-84.

Florida Department of Education. Procedures for Conducting Holistic Scoring for the Essay Portion of the College-Level Academic Skills Test. 1980.

Florida Department of Education. Test Search and Screen for College-Level Communication and Computation Skills. 1981.

Florida Department of Education. Test-Retest Study of the Reliability of the College-Level Academic Skills Test. 1984.

Hambleton, R. K., and Novick, M. R. Toward an integration of theory and method for criterion-referenced tests. Journal of Educational Measurement, 1973, 10, 159-70.

Rasch, G. Probabilistic Models for Some Intelligence and Attainment Tests. 1960. Reprint. University of Chicago Press, 1980.

Ryan, J. ABIL-EST. (Personal communication with J. Ryan of the University of South Carolina, 1981).

Ryan, J. Equating New Test Forms to an Existing Test. Paper at the Annual Meeting of the National Council of Measurement in Education, Los Angeles, 1981.

Stanley, J. C. Reliability. In R. L. Thorndike (Ed.), Educational Measurement (2nd. ed.) Washington, DC: American Council on Education, 1971, 356-482.

Wright, B. D., Mead, R. J., and Bell, S. R. BICAL: Calibrating Items With the Rasch Model. Research Memorandum No. 23C, Statistical Laboratory, Department of Education, University of Chicago, 1980.

41

BEST COPY AVAILABLE

APPENDIX A
CLAST Skills Tested, 1989-90

Essay

Select a topic which lends itself to development.
Determine the purpose and the audience for writing.
Limit the subject to a topic which can be developed within the requirements of time, purpose, and audience.
Formulate a thesis or main idea statement which reflects the purpose and the focus.
Develop the thesis by:
- Providing adequate support which reflects the ability to distinguish between generalized and concrete evidence,
- Arranging the ideas and supporting details in an organizational pattern appropriate to the purpose and focus,
- Writing unified prose in which all supporting material is relevant to the thesis or main idea statement, and
- Writing coherent prose, providing effective transitional devices which clearly reflect the organizational pattern and the relationships of the parts.
- Avoid inappropriate use of slang, jargon, cliches, and pretentious expressions.
- Use a variety of sentence patterns.
- Avoid unnecessary use of passive construction.
- Maintain a consistent point of view.
- Revise, edit, and proofread units of discourse to assure clarity, consistency, and conformity to the conventions of standard American English.

English Language Skills

Word Choice

Use words which convey the denotative and connotative meanings required by context.
Avoid wordiness.

Sentence Structure

Place modifiers correctly.
Coordinate and subordinate sentence elements according to their relative importance.
Use parallel expressions for parallel ideas.
Avoid fragments, comma splices, and fused sentences.

Grammar, Spelling, Capitalization, and Punctuation

Use standard verb forms.
Maintain agreement between subject and verb, pronoun and antecedent.
Use proper case forms.
Use adjectives and adverbs correctly.
Use standard practice for spelling, punctuation, and capitalization.

Reading

Literal Comprehension

Recognize main ideas.
Identify supporting details.
Determine the meanings of words on the basis of context.

Critical Comprehension

Recognize the author's purpose.
Identify author's overall organizational pattern.
Distinguish between statement of fact and statement of opinion.
Detect bias.
Recognize author's tone.
Recognize explicit and implicit relationships within sentences.
Recognize explicit and implicit relationships between sentences.
Recognize valid arguments.
Draw logical inferences and conclusions.

Mathematics

Arithmetic

Add and subtract rational numbers.
Multiply and divide rational numbers.
Add and subtract rational numbers in decimal form.
Multiply and divide rational numbers in decimal form.
Calculate percent increase and percent decrease.
Recognize the meaning of exponents.
Recognize the role of the base number in determining place value in the base-ten numeration system and in systems that are patterned after it.
Identify equivalent forms of positive rational numbers involving decimals, percents, and fractions.
Determine the order-relation between magnitudes.
Identify a reasonable estimate of sum, average, or product of numbers.
Infer relations between numbers in general by examining particular number pairs.
Select applicable properties for performing arithmetic calculations.
Solve real-world problems which do not require the use of variables and which do not involve percent.
Solve real-world problems which do not require the use of variables and which do require the use of percent.
Solve problems that involve the structure and logic of arithmetic.

Algebra

Add and subtract real numbers.
Multiply and divide real numbers.
Apply the order-of-operations agreement to computations involving numbers and variables.
Use scientific notation in calculations involving very large or very small measurements.

Solve linear equations and inequalities.
Use given formulas to compute results when geometric measurements are not involved.
Find particular values of a function.
Factor a quadratic expression.
Find the roots of a quadratic equation.
Recognize and use properties of operations.
Determine whether a particular number is among the solutions of a given equation or inequality.
Recognize statements and conditions of proportionality and variation.
Recognize regions of the coordinate plane which correspond to specific conditions.
Infer simple relations among variables.
Select applicable properties for solving equations and inequalities.
Solve real-world problems involving the use of variables, aside from commonly used geometric formulas.
Solve problems that involve the structure and logic of algebra.

Geometry and Measurement

Round measurements to the nearest given unit of the measuring device.
Calculate distances, areas, and volumes.
Identify relationships between angle measures.
Classify simple plane figures by recognizing their properties.
Recognize similar triangles and their properties.
Identify appropriate types of measurement of geometric objects.
Infer formulas for measuring geometric figures.
Select applicable formulas for computing measures of geometric figures.
Solve real-world problems involving perimeters, areas, and volumes of geometric figures.
Solve real-world problems involving the Pythagorean property.

Logical Reasoning

Deduce facts of set-inclusion or set non-inclusion from a diagram.
Identify simple and compound statements and their negations.
Determine equivalence or nonequivalence of statements.
Draw logical conclusions from data.
Recognize that an argument may not be valid even though its conclusion is true.
Distinguish fallacious arguments from nonfallacious ones.
Infer valid reasoning patterns and express them with variables.
Select applicable rules for transforming statements without affecting their meaning.
Draw logical conclusions when facts warrant them.

Statistics, Including Probability

Identify information contained in bar, line, and circle graphs.
Determine the mean, median, and mode of a set of numbers.
Count subsets of a given set.
Recognize the normal curve and its properties.
Recognize properties and interrelationships among the mean, median, and mode in a variety of distributions.

BEST COPY AVAILABLE

Choose the most appropriate procedures for selecting an unbiased sample from a target population.

Identify the probability of a specific outcome in an experiment.

Infer relations and make accurate predictions from studying particular cases.

Solve real-world problems involving the normal curve.

Solve real-world problems involving probabilities.

APPENDIX B
College-Level Academic Skills Project (CLASP)
and State-Level Task Force Members, 1989-90

CLASP MEMBERS

Project Director
Linda Lou Cleveland, Chipola Junior College

Project Staff
June Siemon, Department of Education
Christy Meeks, Department of Education

Technical Support Contractor (TSC)
Jeaninne N. Webb, Director
Office of Instructional Resources, University of Florida

Standing Committee on Student Achievement
Robert Stakenas, Chairperson, Florida State University
David Alfonso, Palm Beach Community College
Linda Adair, Gulf Coast Community College
R. Scott Baldwin, University of Miami
Richard Burnette, Florida Southern College
Jane Chaney, Brevard County Schools
Elizabeth Cobb, Florida Community College at Jacksonville
Ruth Handley, Superintendent of Highlands County Schools
E. Garth Jenkins, Stetson University
Lola Kerlin, Florida Atlantic University
Robin Largue, Pine Forest High School
John Losak, Miami-Dade Community College
Levester Tubbs, University of Central Florida

COMMUNICATION TASK FORCE MEMBERS

Elizabeth Metzger, Chairperson, University of South Florida
Wilhelmina Boysen, J. M. Tate High School
Joanna Cocchiarella, Satellite High School
Robert Fitzgerald, South Florida Community College
Ann Higgins, Gulf Coast Community College
Jerre Kennedy, Brevard Community College
Gladys Lang, Florida A & M University
Richard Levine, Broward Community College
Jose Marques, Florida International University
Beth Novinger, Tallahassee Community College
Alina Rodriguez, Miami Edison High School
Roy Singleton, University of North Florida
Phillip Taylor, University of Central Florida
Donald Tighe, Valencia Community College

MATHEMATICS TASK FORCE MEMBERS

Charles Goodall, Chairperson, Florida College
Linda Lou Cleveland, Chipola Junior College
Michael Flanagan, Columbia High School
Corinne Garrett, Riverview High School
George Green, Flagler College
Charlene Kincaid, Gulf Breeze High School
Leonard Lipkin, University of North Florida
Alan Mabe, Florida State University
Charles Nelson, University of Florida
Theodore Nicholson, Bethune-Cookman College
Ray Phillips, University of South Florida
Robert Sharpton, Miami-Dade Community College
Karen Walsh, Broward Community College

APPENDIX C
Item Review Guidelines

OVERALL FACTORS TO CONSIDER IN CRITIQUING ITEMS

1. Adequate measurement of skill

2. Fairness of items--items should be free of racial, ethnic, sexual, regional and cultural bias.

3. Quality of stimulus materials (paragraph, graphics, or other material to which students react)--content should be
   a. pertinent and appropriate to grade level;
   b. clear and understandable;
   c. believable and realistic; and
   d. familiar to students of all racial/ethnic backgrounds.

4. Quality of answer choice--there should be
   a. one and only one correct answer, neither too obvious and easy nor too difficult and obscure; and
   b. good distractors, neither too obviously incorrect nor too closely related to the correct answer.

5. Readability of items and instructions--readability should follow guidelines set forth in the test item specifications.

6. Quality of language. The language used should be
   a. clear and concise;
   b. appropriate for grade level;
   c. appropriate for students of all racial/ethnic backgrounds; and
   d. neither too formal and stilted nor too informal and colloquial.

7. Technical considerations--items should be free from flaws such as
   a. too much variation in length of response options;
   b. clues in stem which point to the correct answer;
   c. unclear wording of stem or directions;
   d. confusing use of negative words in stem; and
   e. asking student to choose the correct answer when best answer is really called for (as in choosing the best inference, or the evidence which best supports a given inference), or vice versa.

QUESTIONS TO CONSIDER IN CRITIQUING ITEM CONSTRUCTION

1. Stimulus/stem
   a. Does the stem provide ALL THE INFORMATION necessary to answer the question?
   b. Is the desired response evident by reading the stem alone?
   c. Is the stem written in the POSITIVE (avoids <u>not</u>, <u>except</u>, etc.)?
   d. Is the stimulus portion of the item consistent with the <u>Stimulus Attributes</u>?

2. Response options
    a. Are there four options, arranged in a LOGICAL ORDER (e.g., numerical, alphabetical, chronological)?
    b. Are the options grammatically and conceptually PARALLEL?
    c. Do the options AGREE grammatically with the stem?
    d. Are the options similar and appropriate in LENGTH?
    e. Do the options embody COMMON ERRORS and are they PLAUSIBLE?
    f. Do the options AVOID "all of the above" or "none of the above"?

3. The entire item
    a. Does the item avoid tricky words, phrases, and constructions?
    b. Is the item free of superfluous material and awkward wording?
    c. Does the item avoid unnecessary clues?
    d. Does the item focus on IMPORTANT aspects of content, not trivia?

CONSIDERATIONS IN CRITIQUING ITEMS FOR BIAS

An item is considered to be biased if it contains any language or vocabulary that could benefit or hinder any group's performance. When reviewing an item for bias, one must consider all of the following types of groups of people:

females	regional groups within the U.S.
males	international groups
racial/ethnic groups	religious groups
cultural groups	visually impaired
age groups	hearing impaired
socio-economic groups	other handicaps

As you review each item, consider each of the following questions:

1. Does the item contain any information that could seem to be offensive to any group?

2. Does the item include or imply any stereotypic depiction of any group?

3. Does the item portray any group as degraded in any way?

4. Does the item contain any group-specific language or vocabulary (e.g., culture-related expressions, slang, or expressions) that may be unfamiliar to particular examinees?

APPENDIX D
CLAST Item Specifications Review Team

Project Director
Dianne Buhr

Project Staff
Sue M. Legg
Jeaninne Webb

Reading
Jerre Kennedy, Chairperson, Brevard Community College
Helen Dayan, Hillsborough Community College
Nancy Smith, Florida Community College

English Language Skills
Beth Novinger, Chairperson, Tallahassee Community College
Charles Croghan, Indian River Community College
Elizabeth Metzger, University of South Florida
Betty Owen, Broward Community College
Vincent Puma, Flagler College

Mathematics
Charles Nelson, Chairperson, University of Florida
Nicholas Belloit, Florida Community College
Roy Bolduc, University of Florida
George Coutros, University of Florida
Dennis Clayton, Bethune-Cookman College
Rose Dana, Lake City Community College
Michael Flanagan, Lake City Community College
Leonard J. Lipkin, University of North Florida
Ted Nicholson, Bethune-Cookman College

APPENDIX E

CLAST Item Specifications External Review Committee

Project Director

Linda Lou Cleveland, Chipola Junior College

Project Staff—Department of Education

Thomas H. Fisher, Administrator
Sue Early
Christy Meeks
June Siemon
Dianne Wilkes

Committee Members

Carol Allen, Pasco-Hernando Community College
Faiz Al-Rubaee, University of North Florida
Osiefield Anderson, Florida A and M University
Nancy Brannen, Lake City Community College
Henri Sue Bynum, Indian River Community College
Lynn Cade, Pensacola Junior College
Maureen Cavallaro, Palm Beach Community College
Dale Craft, South Florida Community College
Cathy Denney, St. Johns River Community College
Wayne Dickson, Stetson University
Eunice Everitt, Seminole Community College
Diana Fernandez, Hillsborough Community College
Carl Gabriel, Florida Keys Community College
Barbara Gribble, Gulf Coast Community College
Dorothy Harris, Okaloosa Walton Community College
Bertilda Henderson, Broward Community College
William Kearney, Flagler College
Noel Mawer, Edward Waters College
James Middlebrooks, Edison Community College
Shirley Myers, Florida Community College at Jacksonville
Georgia Newman, Polk Community College
Ron Newman, University of Miami
Cary Ser, Miami-Dade Community College
Barbara Sloan, Santa Fe Community College
Karen Swick, Palm Beach Atlantic College
June White, St. Petersburg Community College
Nora Woodard, Valencia Community College
Raymond F. Woods, Manatee Community College

APPENDIX F
Test-Retest Reliability of the CLAST

In 1984, the Department of Education contracted with Dr. F. J. King of the Florida State University to study certain aspects of the reliability of the College-Level Academic Skills Test (CLAST). Dr. King prepared a report entitled "A Test-Retest Study of the Reliability of the College-Level Academic Skills Test." The study is available from the Department of Education and is summarized herein.

Dr. King invited 360 students who had taken the CLAST in September 1984 to take the CLAST examination a second time. Two hundred seventy-four agreed to do so, and 220 usable scores were obtained. The students were retested in October 1984 with the same form of the test which had been administered in June 1984.

The data were analyzed using several statistics. A Hambleton-Novick (1973) index was calculated to obtain an estimate of the decision consistency over two test forms. The Brennan-Kane (1977) index was used to obtain an index of decision consistency for a single test administration. The KR-20 (Stanley, 1971) index was also calculated because it is a reliability coefficient widely used with norm-referenced tests.

The Hambleton-Novick index calculated with the 1984 passing criteria resulted in the following:

Computation	0.97
Reading	0.86
Writing	0.96
Essay	0.86

The Brennan-Kane indices for the subtests were as follows:

Computation	0.96
Reading	0.96
Writing	0.92
Essay	not applicable

The KR-20 internal consistency coefficients for the subtests resulted in values of:

Computation	0.83
Reading	0.87
Writing	0.74
Essay	not applicable

The reliability coefficients varied depending on which test administration was being analyzed, on the relative difficulty of the tests, and the psychometric characteristics of the tests themselves. Further, it must be recognized that the reported reliability coefficients will vary for subpopulations (e.g., Hispanic) and will vary depending on the placement of the passing criterion.

State of Florida
Department of Education
Tallahassee, Florida
Betty Castor, Commissioner
Affirmative action/equal opportunity employer

124-022291-500-CM-1

BEST COPY AVAILABLE